Perceptual tuning through contact?

Contact interacts with perceptual (not memory-based) face-processing ability to predict cross-race recognition

Joshua Correll
University of Colorado Boulder

Debbie S. Ma
California State University Northridge

Josh P. Davis
University of Greenwich

Abstract

Perceivers generally exhibit better face processing with same-race rather than cross-race faces. To what extent is this deficit attenuated by a perceiver's ability to process faces, and to what extent does that face-processing ability need to be "tuned" by experience with cross-race faces? The current study examined the cross-race recognition deficit (CRD) as a function of participants' ability with faces (measured by one task that emphasizes memory-based ability and one task that measures perceptual ability) and cross-race contact. Our primary analyses involve 583 White participants, 45 of whom can be classified as "super-recognizers." Results suggest that (a) participants with better memory-based face-processing ability generally show a reduction in the CRD, and (b) participants with better perceptual ability only show a reduction in the CRD if they also have cross-race contact. The latter effect suggests that perceptual face processing must be tuned through experience.

In 2017, the New York State Court of Appeals in the United States issued a decision mandating that jurors be educated about the unreliable nature of cross-race eyewitness identification in all cases involving such a witness. The court's rationale relied on scientific evidence of the cross-race recognition deficit (CRD[1]) – the relative difficulty people have in recognizing individuals from racial groups other than their own (Meissner & Brigham, 2001; Young, Hugenberg, Bernstein, & Sacco, 2012). In a typical test of the CRD, participants view a series of randomly presented same-race and cross-race faces in an encoding phase. After a brief delay, they are shown those faces again, intermixed with novel faces of both racial groups, and the participants are asked to identify the faces they saw during the encoding phase, differentiating them from the novel faces. Recognition accuracy is typically quite high for same-race faces, but it drops dramatically for cross-race faces. The CRD is a robust and widely-documented psychological phenomenon that has been replicated across numerous labs, populations, and methods (e.g., Bernstein, Young, & Hugenberg, 2007; Chance, Turner, & Goldstein, 1982; Correll, Lemoine, & Ma, 2011; Pezdek, Blandon-Gitlin, & Moore, 2003; Sangrigoli & De Schonen, 2004).

The current work examines the degree to which the CRD depends on a participant's ability to process, encode, and recognize faces. Quite a bit of research has been conducted investigating individual differences in face processing. This work often shows that different measures of face processing correlate substantially. For example, Verhallen and colleagues (Verhallen, Bosten, Goodbourn, Lawrance-Owen, Bargary, & Mollon, 2017) administered a variety of face-processing measures to a large sample. The authors observed correlations between several tasks in the range of $r = .50$ and argued that this shared variance reflects a general face-processing ability factor. All the same, Verhallen and colleagues also recognize that each of their measures is partially distinct, raising the possibility that each assesses task-specific abilities or subprocesses, such as face recognition versus discrimination (Verhallen et al., 2017, see also Bruce & Young, 1986; and Wilhelm, Herzmann, Kunina, Danthiir, Schacht, & Sommer, 2010, who argue for a distinction between face memory and face perception).

The present work uses two instruments designed to measure different forms of unfamiliar face-processing ability: the Cambridge Face Memory Test: Extended (CFMT+, Russell, Duchaine, & Nakayama, 2009), which is an extended version of the standard CFMT (Duchaine & Nakayama, 2006), and the Glasgow Face Matching Test (GFMT; Burton, White, & McNeil, 2010). Recent work suggests that these tasks typically correlate at about .50 (Balsdon, Summersby, Kemp, & White, 2018; McCaffery, Robertson, Young, & Burton, 2018; Verhallen et al., 2017), but the two measures differ in potentially important ways. The CFMT+ requires the participant to view and encode six faces, hold those faces in memory, and subsequently identify each face from a set of lures. The task thus relies heavily on learning and memory. By contrast, the GFMT presents two faces simultaneously. The photographs are taken at different points in time and with different cameras, and the participant must determine whether the two images depict the same person. Because both images are visible and can be compared in the moment, memory is not involved to the same extent as the CFMT. As such, the GFMT relies more directly on perceptual processes. Both tests measure face-processing ability using White face stimuli only.

Despite scores on these measures correlating in the population (e.g. McCaffery, 2018), there is evidence for dissociations between the underlying processes that each test measures. At the lowest end of the ability spectrum, some individuals with prosopagnosia, or an inability to recognize familiar faces, show impairment on both face memory and face identity perception (i.e. simultaneous face matching), whereas others are impaired on face memory tasks alone (e.g., Bate & Bennetts, 2014; Dalrymple, Garrido, & Duchaine, 2014; for a review see Bate & Bennetts, 2015). At the other end of the spectrum are *super-recognizers*, who possess exceptional *unfamiliar* face learning and recognition abilities (e.g. see Russell et al., 2009). Although most super-recognizers are outstanding at both face memory and face matching tasks, some display relatively poor simultaneous face matching skills (e.g. Bobak, Bennetts, Parris, Jansari, & Bate, 2016; Bobak, Dowsett & Bate, 2016; Bobak, Hancock, & Bate, 2016), while others may be superior at face matching only (e.g. Bate et al., 2018). These dissociations provide some evidence that these tests can partly distinguish between processes driving memory for faces (CFMT+) and those driving perception of faces (GFMT).

When hypothesizing about how face-processing ability may relate to the CRD, we present three possible outcomes. First, it seems possible that people who are more proficient with faces will show a reduction in the magnitude of the CRD. Intuitively, this is an appealing hypothesis – people who are more skilled at recognizing and comparing faces may be less susceptible to cues, like race, that often impair performance for those with lower ability. A second possibility is that general face-processing ability is unrelated to the CRD. This possibility finds some support in the literature. Wilhelm and colleagues find that face-processing ability is domain specific, correlating weakly with verbal and visual recognition (Wilhelm et al., 2010; see also Dennett, McKone, Tavashmi, Hall, Pidcock, Edwards, & Duchaine, 2012). Their work suggests that better face recognition does not translate to improved performance on all types of visual recognition tasks. If individuals who possess superior face-processing ability show the CRD, it would suggest that recognition and matching ability have little impact on the race-based processes that compromise cross-race recognition. Second, a recent paper by Bate and colleagues

(2019) tested the relationship between face-processing ability and the CRD by comparing White and Asian participants with a sample of White super-recognizers (n = 8). The researchers observed no evidence that recognition ability moderated the CRD: super-recognizers exhibited a CRD commensurate with normal participants. Relatedly, Robertson, Black, Chamberlain, Megreya, and Davis (2020) tested a larger sample of white super-recognizers (n = 35) who outperformed White controls at the White face GFMT and a similarly structured Egyptian Face Matching Test. Critically, as with Bate et al. (2019), the strength of the CRD was roughly equal in both groups. Finally, we acknowledge a third possibility, which is of particular interest in the current study. Face-processing ability may reduce the CRD only when participants have the requisite experience with cross-race faces. That is, the cognitive or perceptual tools that allow perceivers to process and remember faces may be necessary, but the beneficial effects of those tools may only emerge if perceivers have had sufficient experience with faces from a specific racial category that enabled perceivers to tune to special patterns of variation among that category of faces. This idea derives from research on *perceptual expertise*, which suggests that, through repeated interaction, participants become sensitized to important patterns of variation in their environment (Rhodes, Ewing, Hayward, Maurer, Mondloch, & Tanaka, 2009; Tanaka & Curran, 2001). Just as novice birders become experts by spending time in the field, perceivers (even those who have the ability) may become experts with cross-race faces only through interaction. We were thus interested in the possibility that face-processing ability interacts with cross-race contact to predict the CRD. Further, such an interaction might also account for the tepid effects of contact on the CRD, when studies fail to account for ability (e.g., Meissner & Brigham, 2001).

Current Research

This study tests two primary questions, which we will address in each analysis as Question One and Question Two. Question One examines whether participants who demonstrate greater face-processing ability show a reduction in the CRD. Again, it is possible that those who are better at processing faces are less susceptible to the influence of race. It is also possible that even participants who are very good at processing faces will be more accurate with same-race faces than with cross-race faces. Question Two examines whether the relationship between face-processing ability and the CRD depends on cross-race contact. Face-processing ability may reduce the CRD more effectively for participants who interact with members of the outgroup. Critically, this study also allows us to examine both questions in terms of (a) general ability, (b) memory-intensive ability (as assessed by the CFMT+), and (c) perceptual ability (as assessed by the GFMT) (cf. Wilhelm et al., 2010).

Methods

All measures, manipulations, and exclusions in the study are reported below, as is the method of determining the final sample size. The CFMT+ and GFMT were collected in advance for a large group of users. Data collection for the current study (i.e., the cross-race recognition task and the contact scale) was conducted in a single uninterrupted wave and was not continued after data analysis. All measures were collected online.

Materials

Cambridge Face Memory Test: Extended (CFMT+) (Russell et al., 2009): The CFMT+ is an extended 102-trial four-stage version of the standardized 72-trial CFMT (Duchaine & Nakayama, 2006). In the first stage, participants are familiarized with images of six White male faces in greyscale, with external facial features cropped (i.e. hairstyle). Subsequently, in each test trial, participants view one of the six faces along with two lures, and make a three-alternative forced-choice decision. Stages become increasingly difficult, and the final 30-trial extended section contains heavily degraded images with larger variations in facial expressions and viewpoints and increases in distractor repetitions. This stage is designed to better discriminate between good and exceptional participants. The CFMT+ is commonly used to assign participants to "super-recognizer" groups. Based on a representative White participant sample ($n = 254$), Bobak (Bobak, Pampoulov, & Bate, 2016) suggest that the super-recognizer threshold should be 95/102 or 2 SD above the mean ($M = 70.72$, $SD = 12.32$). This standard should be achieved by approximately 2% of the population.

Glasgow Face Matching Test (GFMT) (short version: Burton et al., 2010). This self-paced, face-matching test consists of 40 pairs of simultaneously presented male and female White faces in greyscale. Half the trials are matched (i.e., two different photos of the same person), and half mismatched (i.e., two different photos of different persons). Participants respond 'same' if they believe the photos to be a match, or 'different' for mismatches. Burton and colleagues published normalized test scores ($n = 194$; $M = 81.3\%$ [raw score of 32.48], $SD = 9.7$). A maximum score of 40 out of 40 was required to designate a participant as a super-recognizer.

Cross-race face recognition task: We selected eighty male faces (40 Black, 40 White) from the Chicago Face Database (Ma, Correll, & Wittenbrink, 2016). The images were rated by a large number of judges (n = 1,087), allowing us to select faces that were consensually classified as Black or White by at least 90% of raters. Participants completed the cross-race face recognition task on the Qualtrics survey platform. Participants were told that the study was designed to examine memory for faces. They were informed that they would view a series of faces and that they should try to remember them for a subsequent memory test. They then viewed a set of 20 White and 20 Black to-be-remembered faces (selected quasi randomly for each participant) in a mixed order. Each face was presented for 3 seconds before a new face appeared. After completing the encoding phase, participants answered a number of demographic questions, creating a short delay before the memory test. They then viewed the complete set of 80 faces (40 of which had been presented at encoding, 40 of which were novel) in a random order for each participant. For each face, participants indicated whether or not they had seen the face during the encoding phase and their level of confidence in that judgment by responding on a 6-point scale with the following anchors: 1: definitely no; 2: probably no, 3: maybe no, 4: maybe yes, 5: probably yes, 6: definitely yes. This kind of confidence rating is often used to augment simple yes/no judgments, allowing more nuanced signal detection analysis (e.g., identifying multiple criteria or estimating receiver operating characteristic or ROC curves, Macmillan & Creelman, 2004).

Cross-race contact scale (Hancock & Rhodes, 2008). After completing the cross-race recognition task, participants completed a measure of cross-race contact. Eight items of this 15-item questionnaire assess the extent of contact with Black people in a variety of settings, for example, "I live, or have lived, in an area where I interact with African American people." Seven additional items measured the extent of contact with White people. To compute an individual score for cross-race contact, we followed the original procedure outlined by Hancock and Rhodes (2008). We averaged the eight items referring to contact with Black people (alpha=.90) and the seven items referring to contact with White people (alpha=.90). Reliabilities did not differ appreciably as a function of participant race. We then computed a difference score such that positive scores corresponded to more extensive contact with Black people relative to contact with Whites, whereas negative scores corresponded to less extensive contact with Black people relative to contact with Whites.

Participants

Participants completed the CFMT+ and GFMT online prior to the current study and provided contact details for invitation to future studies. From a database of about 40,000 volunteers, all Black participants and a random selection of about 10% of White participants were invited to participate in the current research and asked for consent to access their previous scores on the CFMT+ and GFMT. This approach was based on the idea that, even with low response rates, we would have power to detect small effects.

The stimuli were Black and White faces, so our predictions apply to monoracial Black and White participants because these participants view one clear racial ingroup and one clear racial outgroup. Other racial groups and participants reporting multi-racial identity (n=3) were not analyzed. Because we had limited participation from Black participants, the primary analyses focus on 583 individuals identifying as White (369 female, 211 male, 3 other; mean age = 36.08, [SD = 11.00]). A sensitivity analysis indicates that a sample of this size will detect effects as small as η^2=.0133 with power of .80. These participants were characterized by somewhat better-than-average face processing, though the range of performance was substantial, M_{CFMT+}=86.23 (SD=9.39, min=43, max=101), M_{GFMT}=37.36 (SD=2.45, min=26, max=40). These participants were also characterized by greater contact with White people, M_{White}=4.91 (SD=1.28), than Black people, M_{Black}=2.20 (SD=1.04), and this difference was significant, $t(582)=39.42$, $p<.001$.

We will present additional analyses of 46 participants who achieved super-recognizer status on both measures, as described above (1 Black female, 29 White females, 16 White males). These individuals are included in the primary analyses of White and Black participants, but we will also consider them separately at the end of the results section. Because of ethical and statistical complications involved in analyzing a single member of any category, all super-recognizers tests will focus on the 45 White super-recognizers (mean age = 35.09). These participants were necessarily characterized by extremely high levels of face processing with little variability, M_{CFMT+}=97.09 (SD=1.77, min=95, max=101), M_{GFMT}=40 (SD=0). Like White participants in general, the White super-recognizers were characterized by greater contact with White people, M_{White}=4.76 (SD=1.42), than Black people, M_{Black}=2.25 (SD=1.00), and the difference was significant, $t(44)=9.83$, $p<.001$.

Finally, we will present analyses of the 40 participants who identified as Black (19 female, 21 male; mean age = 33.98 [SD=10.00]). These participants were also characterized by better-than-average face processing, and again the range was substantial, M_{CFMT+}=82.75 (SD=12.29, min=48, max=100), M_{GFMT}=35.95 (SD=2.45, min=26, max=40). Somewhat surprisingly, these participants were characterized by greater contact with White people (a racial outgroup), M_{White}=4.23 (SD=1.51), than Black people (the ingroup), M_{Black}=3.83 (SD=1.53), but the difference was smaller than the corresponding difference among Whites, and did not reach significance, $t(39)=1.35$, $p=.184$. Full analyses of both subsamples (Black participants and super-recognizers) are available in the supplemental materials.

Design

The study involved a standard measure of the CRD in which participants view several to-be-remembered faces and then are shown a set of test faces, including all of the to-be-remembered faces and an equal number of lures. The design involved two within-participant factors: prior presentation (was the test face viewed during the encoding phase or not?) and face race (was the test face White or Black?). Between participants, we measured cross-race contact and face-processing ability using both the CFMT+ and GFMT.

Results

Analytic Strategy

The data were submitted to a series of linear mixed-model analyses treating both participants and stimuli as random factors (Judd, Westfall, & Kenny, 2012). Recall that each participant viewed 80 faces during the test phase (40 Black, 40 White). Half of these faces (20 Black, 20 White) had been presented during encoding, and half were novel. Participants rated each test face on a 6-point scale of increasing confidence that the face had been viewed during the encoding phase. The main analyses, reported below, examined the full variability of this measure. We also conducted modified analyses, first recoding the six-point scale into binary judgments about whether the participant believed she or he saw the face during encoding (1-3 = "did not see", 4-6 = "did see") (see supplementary material). When applied to this kind of binary data, a generalized linear mixed model with a probit link function is essentially equivalent to a signal detection analysis (DeCarlo, 1998). Second, we conducted an ordinal analysis because the data clearly suggest that variation among the faces that were presented at encoding exceeds the variation of the not-seen faces[2]. This pattern is common, but it may compromise the primary analysis (which assumes interval scaling). By relaxing that assumption, ordinal analysis should shed light on the robustness of the reported effects. Although these alternative approaches ignore potentially important variation in recognition ratings, they yield very similar results. The only result that emerges in the primary analysis but not in the ordinal/binary analyses involves the relationship between the CFMT+ (in Analysis 2) and the CRD. Critically, all effects of contact replicate across analyses.

We analyzed the ratings as a function of whether the face had actually been presented during the encoding phase or not (prior presentation: presented = +1; not presented = -1), whether the face was White or Black (face race: White = +1; Black = -1), the participant's self-reported cross-race contact (measured continuously and mean-centered), face-processing ability as measured by the CFMT+ and/or GFMT (measured continuously and mean-centered), and all higher-order interactions. Across participants, the models allowed for random effects of prior presentation, face race, and the prior presentation × face race interaction. Across faces, the models allowed for random effects of prior presentation. The models were estimated using the lme4 package in R, allowing an unstructured variance-covariance matrix. Though effect sizes are not well defined for models with crossed random factors, we report 95% bootstrapped confidence intervals for each test. The data, R code, and results are available online (https://osf.io/wbgzj/?view_only=39d7923282794b1aa7a0734624926143). From this point on, results refer to only White participants except where participant race is explicitly mentioned.

In this sample, as in previous work (McCaffery et al., 2018), the CFMT+ and GFMT were correlated, $r(581)=.518$, $p<.001$. The correlation indicates that roughly 25% of the variance in each measure is shared with the other. This common variation presumably reflects general face-processing ability (Verhallen et al., 2017). By the same logic, this correlation also indicates that roughly 75% of the variation in each measure is unshared, reflecting either error or (more interestingly) *task-specific individual differences*, such as memory or perceptual ability (Wilhelm et al., 2010). Accordingly, we explored the common and idiosyncratic effects of these measures. Our first analysis tested the common variance by standardizing and averaging the two measures to form an index of general face-processing ability. In our second analysis, we entered both measures as simultaneous predictors in the model. In this analysis, we thus ignore the common variance, and instead test the partial or unique effects of memory-based (CFMT+) and perceptual (GFMT) ability. For each measure of face-processing ability (general, uniquely memory-based, and uniquely perceptual), we examine our two key questions. Question One asks, do participants with better face-processing ability show a reduction in the CRD? Question Two asks, does the relationship between face-processing ability and the CRD depend on cross-race contact?

Effects of general face-processing ability (Analysis 1)

This analysis examines recognition confidence as a function of prior presentation, face race, cross-race contact, and *general* face-processing ability (the average of the standardized CFMT and GFMT scores). The analysis was performed using a mixed-effects model, as described above. Before addressing the questions that motivated this study, it is important to understand a number of lower-order effects. First, in this analysis, we observed a pronounced effect of prior presentation, which (in line with signal detection theory) we will refer to as *sensitivity*, *b=1.610, CI=[1.534,1.683],* $t(113.0) = 44.31$, $p < .001$. This effect indicates that participants were more confident that they had seen a face if that face had, in fact, been presented during the encoding phase than if the face had not been presented during the encoding phase (i.e., participants could discriminate between familiar and unfamiliar faces). Second, we replicated the CRD effect (reflected by the interaction of prior presentation and race), such that sensitivity depended on the race of the face, *b=0.180, CI=[0.111,0.244],* $t(79.5) = 5.42$, $p < .001$. Participants were more sensitive (better at differentiating familiar from unfamiliar) for White faces than for Black faces. Third, sensitivity was also moderated by face-processing ability, *b=0.234, CI=[0.198,0.269]*, $t(579.0) = 12.46$, $p < .001$. This interaction is not surprising. It suggests that people who performed well on previous measures of recognition also performed well on the current recognition task. Finally, but somewhat surprisingly, these data offered no evidence of a relationship between cross-race contact and the CRD (this is tested by the contact × prior presentation × race interaction), *b=0.003, CI=[-0.012,0.016]*, $t(577.8) = 0.40$, $p = .69$. This null effect is noteworthy given our large sample and the meta-analytic finding that contact is, generally, related to the CRD. Still, as noted above, the effect of contact is typically small, and non-significant correlations are common. These basic effects provide a context in which we can evaluate the questions that motivated the current work.

Question One, our first critical question, asks, does general face-processing ability moderate the CRD? This question is tested by the three-way interaction between prior presentation, face race, and the participant's general ability score (again, an average of the standardized CFMT and GFMT scores). This interaction was not significant, *b=-0.010, CI=[-0.025,0.005]*, $t(578.7) = -1.31$, $p = .191$. Using a general measure of face-processing ability, we found no evidence that participants who have greater ability show a reduction in the CRD. Our second critical question, Question Two, asks whether the effect of ability depends on contact. There was no evidence that contact moderated the relationship between face-processing ability and the CRD, *b=-0.011, CI=[-0.026,0.003]*, $t(577.1) = -1.35$, $p = .175$. Though both trends were in the predicted direction, there was little evidence that general face-processing ability was related to the CRD. Figure 1 presents ROC curves for both Black and White faces for participants who are higher / lower than average in terms of general ability. Relative to the Black curves, the White curves' more pronounced deviation from the diagonal line reflects the participants' superior sensitivity to White faces. Similarly, the more pronounced curvature for high-ability participants represents their superior sensitivity.

***Unique effects of memory-based and perceptual face-processing ability (Analysis 2)*:**

This analysis examined the ratings as a function of prior presentation, face race, cross-race contact, CFMT+ scores, and (simultaneously) GFMT scores. Again, we used a mixed-effects model, as described above. The two ability measures (CFMT+ and GFMT) were not allowed to interact, but all other interactions were included. As this analysis includes both the CFMT+ and the GFMT, the effects of each measure necessarily partial out the influence of the other (focusing on the 75% of variance that is not shared with the other measure). So, observed effects of the CFMT+ reflect effects of memory-based ability holding constant perceptual ability, and effects of the GFMT reflect effects of perceptual ability holding constant memory-based ability. Results again showed evidence of sensitivity: participants were more confident that they had seen a face if they actually saw that face during encoding, $b=1.610$, $CI=[1.531,1.682]$, $t(112.6) = 44.34$, $p < .001$. As in the previous model, this sensitivity was moderated by both the race of the face, $b=0.179$, $CI=[0.116,0.252]$, $t(79.4) = 5.39$, $p < .001$ (sensitivity was higher for White faces than Black faces), and both measures of face-processing ability, CFMT+: $b=0.166$, $CI=[0.128,0.208]$, $t(576.8) = 8.78$, $p < .001$; GFMT: $b=0.067$, $CI=[0.029,0.108]$, $t(577.3) = 3.52$, $p < .001$ (participants with higher scores on each measure displayed greater sensitivity, even controlling for the other measures). As before, this analysis yielded no evidence that contact reduces the CRD, $b=0.004$, $CI=[-0.009,0.017]$, $t(575.7) = 0.62$, $p = .538$.

Effects of memory-based ability. Next, we explore effects that can be attributed uniquely to memory-based ability, controlling statistically for perceptual ability. Question One involves the idea that, given an average level of cross-race contact, participants with superior memory ability might show a reduction in the CRD. This question is tested by a three-way interaction between prior presentation, race, and CFMT+ scores. This interaction was significant, such that participants with higher CFMT+ scores showed a reduction in the magnitude of the CRD, *b=-0.020, CI=[-0.036,-0.005],* $t(575.3) = -2.49$, $p = .013$. Question Two involves the idea that the effect of memory ability depends on cross-race contact. There was no evidence that the three-way interaction, described above, depended further on cross-race contact, *b=0.012, CI=[-0.003,0.027],* $t(575.8) = 1.46$, $p = .145$. We can characterize the general pattern of CFMT+ effects as follows: participants who have better memory for faces show a reduction in the CRD, but the magnitude of this reduction does not seem to depend on cross-race contact. Figure 2 shows ROC curves for higher and lower levels of performance on the CFMT+. Though the curves for White faces clearly show higher sensitivity (greater curvature) than the curves for Black faces, the discrepancy is reduced for participants with greater ability.

Effects of perceptual ability. In the same model (which includes both CFMT+ and GFMT as simultaneous predictors) we explored effects that can be attributed uniquely to perceptual ability, controlling statistically for memory-based ability. Question One, does perceptual ability moderate the CRD, is tested by a three-way interaction between prior presentation, face race, and the participant's GFMT score. This interaction was not significant, offering no evidence that, given average levels of cross-race contact, participants with better perceptual ability showed a reduction in the CRD, *b=0.011, CI=[-0.007,0.028],* $t(577.5) = 1.34$, $p = .181$. However, considering Question Two, we observed a significant four-way interaction suggesting that the effect of perceptual ability depends on cross-race contact, *b=-0.023, CI=[-0.039,-0.009],* $t(575.7) = -2.95$, $p = .003$. Among participants who rarely interact with Black people (levels of cross-race contact one standard deviation below the mean), increases in perceptual ability were associated with a *more severe* CRD, *b=0.034, CI=[0.011,0.057], t*(576.3) = 2.81, *p* = .005. This effect emerges because, in the absence of cross-race contact, people with greater (rather than lesser) perceptual ability show dramatic improvements in sensitivity to White faces, *b=0.095, CI=[0.032,0.154], t*(577.1) = 3.02, *p* = .003, but no significant improvement in sensitivity to Black faces, *b=0.026, CI=[-0.036,0.090], t*(576.8) = 0.82, *p* = .413. This latter null effect is somewhat remarkable. It suggests that, when contact is low, better perceptual ability does not lead to any observable improvement in recognition of Black faces. In stark contrast, among participants who interact frequently with Black people (levels of contact one standard deviation above the mean), higher GFMT scores were not associated with a meaningful change in the CRD, *b=-0.013, CI=[-0.035,0.007], t*(577.1) = -1.20, *p* = .230. This null result occurred because, with higher levels of cross-race contact, participants with greater perceptual ability showed increases in sensitivity to *both* White faces, *b=0.062, CI=[0.012,0.117], t*(577.3) = 2.33, *p* = .020, and Black faces, *b=0.087, CI=[0.037,0.137], t*(577.1) = 3.24, *p* = .001. This pattern is consistent with the idea that perceptual ability can be "tuned" by experience. For the GFMT, then, the overall pattern suggests that perceptual ability can reduce the CRD, but its effects are contingent on cross-race contact. See Figure 3.

Ancillary Results

Across analyses, there were effects that do not relate to the questions of interest in this study. We report them for the sake of completeness, but we do not seek to interpret them. First, participants with higher average ability and higher CFMT+ scores responded with more confidence that they had seen faces in general (i.e., they used the higher end of the response scale), average: *b=0.045, CI=[0.010,0.083]*, *t*(579.1) = 2.45, *p* = .015; CFMT+: *b=0.058, CI=[0.022,0.093]*, *t*(576.9) = 3.10, *p* = .002. (The corresponding partial effect of GFMT scores was not significant, *b=-0.014, CI=[-0.050,0.021]*, *t*(577.4) = -0.74, *p* = .46.) Second, participants who reported greater cross-race contact demonstrated lower sensitivity, *b=-0.035, CI=[-0.071,-0.004]*, *t*(576.9) = -2.27, *p* = .023 (this pattern emerged in both analyses, but this statistic is taken from the first analysis).

An examination of White super recognizers (Analysis 3)

Another way to examine the effects of face-processing ability on the CRD is to examine super-recognizers – people with very high levels of proficiency on both measures of face-processing. We can test our hypotheses by looking to see whether super-recognizers differ from other participants in terms of either a reduced CRD or a stronger relationship between contact and the CRD. To that end, we identified participants with scores of 95 or higher on the CFMT+ and 40 on the GFMT as super-recognizers. The current dataset included a substantial number who exceeded both thresholds. There were 46 super-recognizers, 45 White and 1 Black. We examined the performance of the White super-recognizers to see if there was evidence of a CRD among this high-performing group. The data were analyzed as a function of prior presentation, the race of the face, and the extent of cross-race contact.[3]

Not surprisingly, super-recognizers demonstrated high levels of sensitivity to prior presentation, *b=1.900*, *CI=[1.793,2.011]*, *t*(69.3) = 32.52, *p* < .001. When rating test faces, they were more confident in having seen a stimulus if, in fact, that stimulus had been presented at encoding. As a check on the superior recognition ability of these individuals, we note that the magnitude of the prior presentation effect among super-recognizers was substantially greater than the corresponding effect in the White sample as a whole (*b* = 1.61), and larger than participants who did not qualify as super-recognizers (*b* = 1.59). A test comparing the magnitude of the prior presentation effect among super-recognizers versus non-super-recognizers was significant, *b=0.156, CI=[0.088,0.223], t*(578.8) = 4.58, *p* < .001. But, in spite of their general proficiency, super-recognizers showed clear, robust evidence of the CRD, *b=0.187, CI=[0.116,0.264], t*(77.9) = 5.19, *p* < .001. In fact, a direct comparison of super-recognizers to non-super-recognizers offers no evidence that the magnitude of the CRD is reduced among super-recognizers, *b=0.002, CI=[-0.025,0.026], t*(578.3) = 0.153, *p* = .88. This null effect of even an extreme level of face-processing ability is consistent with the null effects of continuously measured general face-processing ability reported above (using the average of the CFMT+ and GFMT) and with prior work with super-recognizers (e.g., Bate et al., 2019). See Figure 4. Within the sample of super-recognizers, there was no evidence that contact moderated the CRD, *b=0.015, CI=[-0.024,0.053], t*(3373.9) = 0.786, *p* = .43; and when comparing the super-recognizers to non-super-recognizers, we found no evidence that the magnitude of the contact-CRD relationship (which was not significant, itself) depended on super-recognizer status, *b=0.004, CI=[-0.023,0.032], t*(579.2) = 0.300, *p* = .76.

An examination of the Black participants (Analysis 4)

We analyzed the data from the Black sample using models similar to those used with the White sample to facilitate comparison (see Footnote 2). The results should be viewed with some skepticism because of (a) the small number of participants, (b) the complexity of this analysis, and (c) the fact that the Black participants demonstrated higher levels of contact with White people than with Black people (the opposite of the homophily observed with White participants).

We observed no evidence of a CRD, with sensitivity that was directionally (but not significantly) higher for Black faces than for White faces, *b=-0.093, CI=[-0.214,0.035]*, *t*(192.7) = -1.410, *p* = .160. There was no evidence that variation in face-processing ability (CFMT+ and GFMT scores, combined or uniquely) moderated the CRD. In contrast to the White sample, there was no evidence that participants with better memory-based ability showed a reduction in the CRD, *b=-0.033, CI=[-0.115,0.053], t*(285.4) = -0.76, *p* = .45, nor that the advantages of greater perceptual ability depend on cross-race contact, *b=0.002, CI=[-0.046,0.056]*, *t*(284.4) = 0.06, *p* = .95. Again, these very weak patterns may be due, in part, to the small sample of Black participants, but they may also be due to the fact that (a) Black participants in majority-White countries may have more exposure to White faces (Wan, Crookes, Dawel, Pidcock, Hall, & McKone, 2017) and/or (b) the CFMT+ and GFMT assess face-processing ability primarily using White faces, complicating the interpretation of these measures in a Black sample. See Figure 5. We return to this issue in the Discussion.

Discussion

The current investigation examined variation in the CRD as a function of face-processing ability and cross-race contact. The underlying rationale is that, although face-processing ability should impact recognition of faces, in general, it may have a larger effect if participants have experience with members of other racial groups. Cross-race contact should enable perceivers to learn which cues are diagnostic in cross-race faces in the same manner that they do for same-race faces (Brigham & Barkowitz, 1978; Brigham & Malpass, 1985). The same logic suggests that cross-race contact may help reduce the CRD, but the effects of that contact may depend on the perceiver's overall ability to process faces. Indeed, meta-analytic work suggests that cross-race contact weakly relates to the CRD (Meissner & Brigham, 2001). The current work might help to explain why contact alone may be insufficient for attenuating the CRD by highlighting the moderating role of face-processing ability.

To investigate this question, we administered a cross-race face recognition test and a cross-race contact questionnaire to a large sample of mostly White participants. These participants had previously been assessed on two measures of face-processing ability, one that relies heavily on memory and one that relies heavily on concurrent perceptual processing. We examined whether face-processing ability was related to the CRD, whether that relationship depended on cross-race contact, and whether the memory-intensive and perceptual measures of ability had different effects. The results suggest that general face-processing ability (as measured by the average of the two widely used face measures) does not moderate the CRD. Participants who are especially proficient with faces – even super-recognizers – do not show a meaningful reduction in the CRD (see also Bates et al., 2019; Robertson et al., 2020). The results also suggest different patterns of effects for memory versus perceptual face-processing ability. First, participants with higher scores on the memory-based measure showed reductions in the CRD[4]. Second, the relationship between perceptual ability and the CRD depended on cross-race contact, such that the capacity of perceptual ability to reduce the CRD was stronger among participants with more extensive cross-race contact. Though people certainly seem to vary meaningfully in terms of general ability (Verhallen et al., 2017), these data clearly suggest that face-processing is a multifaceted construct, and that these facets have distinct effects on the processing of cross-race faces.

These data may also offer some important hints about the psychological processes that give rise to the CRD. Available research suggests that motivational, cognitive, and perceptual deficits with cross-race faces may all contribute to the CRD. We briefly explain each of these effects. First, relative to a same-race face, a perceiver may be less motivated to individuate a cross-race face and may therefore exert less effort during encoding or recognition (e.g., Bernstein et al., 2007; Young & Hugenberg, 2011). Second, perceivers may pay greater attention to information about the racial category of a cross-race face than a same-race face (Levin, 1996, 2000). It may seem that more attention to a cross-race face would facilitate recognition; however, perceivers seem to attend to the wrong kinds of visual information, focusing on *category-specifying* information (e.g., the fact that a face belongs to the category Black) instead of individuating information that would help differentiate one cross-race face from others (e.g., the face has wider eyes than other Black faces). Third, because most people live in segregated environments (spending more time with racial ingroup members), a perceptual expertise account suggests that the perceptual system becomes “tuned” to visual information that differentiates same-race faces (Correll, Hudson, Guillermo, & Earls, 2016). Note that this prioritization of diagnostic cues requires experience with category exemplars (Tanaka & Curran, 2001).

In our study, as well as in other studies, this experience is operationalized as cross-race contact (e.g., Chiroro & Valentine, 1995; Rhodes et al., 2009; Sangrioli, Pallier, Argenti, Ventureyra, & Schonen, 2005). The current study was not designed to disentangle these three accounts, but it is interesting to note that the moderating effects of cross-race contact emerged only for the more perceptually oriented GFMT. Perceptual ability is associated with a reduced CRD, but only when participants have higher levels of cross-race contact. Motivational and attention-based accounts of the CRD would seem to predict a different pattern. For example, if cross-race contact increases the motivation of a White participant to individuate Black faces, that motivation should (a) directly reduce the CRD (yielding a “main effect”), and (b) magnify the effects of any kind of ability (yielding interactions with both the CFMT+ and the GFMT). This is not the pattern we observed. Exactly the same predictions emerge if cross-race contact reduces category-focused attention: we should again observe a direct effect of contact and an interaction with both ability measures. The fact that cross-race contact interacts only with the GFMT but not with the CFMT+ suggests that the effects of contact are perceptual in nature. This is not a conclusive test, but it lends some weight to the perceptual expertise account.

On the GFMT, participants make a series of decisions about pairs of novel faces (which are never repeated), and these decisions may involve broad face-to-face comparisons, narrow feature-to-feature comparisons, or a combination of both. Face matching performance can be improved by feature-by-feature strategies (Towler, White, & Kemp, 2017). On the CFMT+, six target faces are repeated throughout the test, displayed from different viewpoints and with different expressions. Accuracy requires the participant to develop rudimentary abstract face representations in memory. The CFMT+ may thus be linked more to memory-based strategies that involve the entire face (e.g. see Tanaka & Farah, 1993), whereas the GFMT may be sensitive to the kinds of perceptual expertise thought to facilitate differentiation of same-race faces (e.g. Correll et al., 2016). Further research could evaluate whether performances on tasks more specifically designed to measure these processes (e.g. part-whole effect; see Tanaka, Heptonstall, & Campbell, 2019) might shed light on the current results.

It should be noted that the CFMT+ and GFMT have been criticized by some researchers (e.g. Bate et al., 2018). For instance, some prosopagnosics may use task-specific strategies to generate high scores on the CFMT (e.g. Esins, Schultz, Stemper, Kennerknecht, & Bulthoff, 2016), while the GFMT may suffer from ceiling effects (Davis et al., 2016). Since carrying out the current study, newer more reliable and valid face memory (e.g. Bate et al., 2018; Dunn, Summersby, Towler, Davis, & White, 2020) and matching (e.g. Fysh & Bindemann, 2017) tests have been developed to address some of these limitations. Research on this topic should incorporate these tests to validate our findings.

A third obvious and important limitation involves the lack of diversity in the sample used in the primary analyses. The hypotheses of this study pertain only to White and Black people, and we have a relatively small sample of Black participants. Accordingly, the primary analyses were conducted with the White sample. This limitation is especially important given the fact that face-processing ability in this study, as in much research, was measured using tasks that employ White faces. Although we have been discussing the CFMT+ and GFMT as measures of general face-processing ability, for our predominantly White sample, higher scores on the CFMT+ and GFMT also reflect better memory and perceptual ability for same-race faces. There is indeed evidence that performance on these tasks depends on perceiver and target race (e.g., Wan et al., 2017). Additional data from a larger and perhaps more representative sample of Black participants and research that manipulates the race of the face stimuli used in the CFMT+ and GFMT would help clarify the effects of race (as distinct from face-processing ability) on task performance.

However, the lack of diversity in our sample cannot easily explain the effects of interest. For example, if the CFMT+ and GFMT reflect *only* ability with same-race faces, we would expect that participants with high scores would demonstrate very good recognition for White faces (and only for White faces) leading to a positive association between each measure and the CRD. In contrast, we observe that the CFMT+ is negatively associated with the CRD. Further, the over-representation of White participants cannot explain the divergence we observe in the effects of the two measures (memory-based versus perceptual), nor can it explain the moderating effect of cross-race contact on the effects of the GFMT. Again, if the GFMT reflects *only* perceptual ability with same-race faces, there is no reason to think that its relationship with the CRD should depend on cross-race contact. But in fact, this interaction seems quite robust. It seems more parsimonious to conclude that the GFMT generally (if imperfectly) assesses perceptual ability, and that perceptual ability facilitates recognition of cross-race faces for participants who interact with members of that racial outgroup. As with many questions in psychology, future work would certainly benefit from more diverse samples, but the lack of diversity, here, cannot easily explain the findings we report.

Research investigating the CRD has broadened our understanding of how the social world shapes the human mind. As a field, researchers have explored fundamental basic science questions surrounding the CRD, and exposed the complex interplay between perception, attention, learning, and motivation. More significantly, these efforts have yielded concrete improvements to people's lives by informing policy and law enforcement. As we described at the outset of this paper, the justice system acknowledges the scientific value of this work when they change procedures to render fairer judgments in our courts. Working toward identifying mechanisms and developing interventions for mitigating the CRD constitute important steps both for scientific discover and for the translation of this work to the real world.

Open Practices

All data, code, and output for the analyses reported above can be found online at: https://osf.io/wbgzj/?view_only=39d7923282794b1aa7a0734624926143

Footnotes

[1] Existing literature uses many different names to describe superior recognition of racial ingroup faces, including the own-race bias, the other-race effect, and the cross-race recognition deficit. We prefer the last term because it is more specific than other-race effect (which is extremely vague) and more accurate than own-race bias (which is misleading, because the effect is not a bias in signal detection terms).

[2] We thank an anonymous reviewer for suggesting this analysis.

[3] We attempted to keep the models of the smaller samples (i.e., the super recognizers on their own and Black participants on their own) consistent with the primary analyses, reported above. However, with these models (which are fairly complex), there were instances of singular fit when estimating the confidence intervals. We were therefore forced to simplify the random effects portions of the models. Even the simplest models (using only random intercepts) generated some singular fit in estimating the confidence intervals, but these problems were rare (less than 4% of bootstrap iterations).

[4] But note that this is the only effect that did not replicate when recognition was treated as a binary or ordinal outcome.

References

Balsdon, T., Summersby, S., Kemp, R. I., & White, D. (2018). Improving face identification with specialist teams. *Cognitive Research: Principles and Implications, 3*(1), 25.

Bate, S., & Bennetts, R. J. (2014). The rehabilitation of face recognition impairments: a critical review and future directions. *Frontiers in Human Neuroscience, 8*(July), 1–17.

Bate, S., & Bennetts, R. (2015). The independence of expression and identity in face processing. Evidence from neuropsychological case studies. *Frontiers in Psychology, 6*, 770.

Bate, S., Bennetts, R., Hasshim, N., Portch, E., Murray, E., Burns, E., & Dudfield, G. (2018). The limits of super recognition: An other-ethnicity effect in individuals with extraordinary face recognition skills. *Journal of Experimental Psychology: Human Perception and Performance, 45*(3), 363-377.

Bate, S., Frowd, C., Bennetts, R., Hasshim, N., Murray, E., Bobak, A. K., … Richards, S. (2018). Applied screening tests for the detection of superior face recognition. *Cognitive Research: Principles and Implications*, *3*(22), 1–19.

Bobak, A. K., Bennetts, R. J., Parris, B. A., Jansari, A., & Bate, S. (2016). An in-depth cognitive examination of individuals with superior face recognition skills. *Cortex, 82*, 48–62.

Bobak, A. K., Dowsett, A., & Bate, S. (2016). Solving the border control problem: Evidence of enhanced face matching in individuals with extraordinary face recognition skills. *PLoS One, 11*, e0148148

Bobak, A. K., Hancock, P. J. B., & Bate, S. (2016). Super-recognizers in action: Evidence from face matching and face memory tasks. *Applied Cognitive Psychology, 30,* 81–91.

Bobak, A. K., Pampoulov, P., & Bate, S. (2016). Detecting superior face recognition skills in a large sample of young British adults. *Frontiers in Psychology*, *7,* 1378.

Bernstein, M. J., Young, S. G., & Hugenberg, K. (2007). The cross-category effect: Mere social categorization is sufficient to elicit an own-group bias in face recognition. *Psychological Science*, *18*(8), 706-712.

Brigham, J. C., & Barkowitz, P. (1978). Do "They all look alike?" The effect of race, sex, experience, and attitudes on the ability to recognize faces. *Journal of Applied Social Psychology*, *8*(4), 306-318.

Brigham, J. C., & Malpass, R. S. (1985). The role of experience and contact in the recognition of faces of own-and other-race persons. *Journal of Social Issues*, *41*(3), 139-155.

Bruce, V., Henderson, Z., Newman, C., & Burton, A. M. (2001). Matching identities of familiar and unfamiliar faces caught on CCTV images. *Journal of Experimental Psychology: Applied*, *7*(3), 207.

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology*, *77*(3), 305-327.

Burton, A. M., White, D., & McNeill, A. (2010). The Glasgow face matching test. *Behavioral Research Methods*, *42*, 286–291.

Caldara, R., & Abdi, H. (2006). Simulating the 'other-race'effect with autoassociative neural networks: further evidence in favor of the face-space model. *Perception*, *35*(5), 659-670.

Caldara, R., Rossion, B., Bovet, P., & Hauert, C. A. (2004). Event-related potentials and time course of the 'other-race'face classification advantage. *Neuroreport*, *15*(5), 905-910.

Chance, J. E., Turner, A. L., & Goldstein, A. G. (1982). Development of differential recognition for own-and other-race faces. *The Journal of Psychology*, *112*(1), 29-37.

Chiroro, P., & Valentine, T. (1995). An investigation of the contact hypothesis of the own-race bias in face recognition. *The Quarterly Journal of Experimental Psychology Section A*, *48*(4), 879-894.

Correll, J., Hudson, S. M., Guillermo, S., & Earls, H. A. (2017). Of kith and kin: Perceptual enrichment, expectancy, and reciprocity in face perception. *Personality and Social Psychology Review*, *21*(4), 336-360.

Correll, J., Lemoine, C., & Ma, D. S. (2011). Hemispheric asymmetry in cross-race face recognition. *Journal of Experimental Social Psychology*, *47*(6), 1162-1166.

Dalrymple, K. A., Garrido, L., & Duchaine, B. (2014). Dissociation between face perception and face memory in adults, but not children, with developmental prosopagnosia. *Developmental Cognitive Neuroscience*, *10*, 10–20.

DeCarlo, L. T. (1998). Signal detection theory and generalized linear models. *Psychological Methods*, *3*(2), 186-205.

Dennett, H. W., McKone, E., Tavashmi, R., Hall, A., Pidcock, M., Edwards, M., & Duchaine, B. (2012). The Cambridge Car Memory Test: A task matched in format to the Cambridge Face Memory Test, with norms, reliability, sex differences, dissociations from face memory, and expertise effects. *Behavior Research Methods*, *44*(2), 587-605.

Duchaine, B., Germine, L., & Nakayama, K. (2007). Family resemblance: Ten family members with prosopagnosia and within-class object agnosia. *Cognitive Neuropsychology*, *24*(4), 419-430.

Duchaine, B., & Nakayama, K. (2006). The Cambridge Face Memory Test: Results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. *Neuropsychologia*, *44*(4), 576-585.

Dunn, J., Summersby, S., Towler, A., Davis, J. P., & White, D. (2020, May 29). UNSW Face Test: A screening tool for super-recognizers. Retrieved from psyarxiv.com/k7mf6

Esins, J., Schultz, J., Stemper, C., Kennerknecht, I., & Bulthoff, I. (2016). Face perception and test reliabilities in congenital prosopagnosia in seven tests. *i-Perception, 7*(1), 1-37.

Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. In *Advances in Experimental Social Psychology* (Vol. 23, pp. 1-74). Academic Press.

Furl, N., Phillips, P. J., & O'Toole, A. J. (2002). Face recognition algorithms and the other-race effect: computational mechanisms for a developmental contact hypothesis. *Cognitive Science*, *26*(6), 797-815.

Fysh, M. C., & Bindemann, M. (2017). The Kent Face Matching Test. *The British Journal of Psychology, 109,* 219-231.

Hancock, K. J., & Rhodes, G. (2008). Contact, configural coding and the other-race effect in face recognition. *British Journal of Psychology*, *99*(1), 45-56.

Hugenberg, K., & Sacco, D. F. (2008). Social categorization and stereotyping: How social categorization biases person perception and face memory. *Social and Personality Psychology Compass*, *2*(2), 1052-1072.

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*, *103*(1), 54.

Levin, D. T. (1996). Classifying faces by race: The structure of face categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(6), 1364.

Levin, D. T. (2000). Race as a visual feature: Using visual search and perceptual discrimination tasks to understand face categories and the cross-race recognition deficit. *Journal of Experimental Psychology: General*, *129*(4), 559.

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods*, *47*(4), 1122-1135.

Macmillan, N. A., & Creelman, C. D. (2004). Detection theory: A user's guide. Psychology press.

McCaffery, J. M., Robertson, D. J., Young, A. W., & Burton, A. M. (2018). Individual differences in face identity processing. *Cognitive research: Principles and implications*, *3*(1), 21.

Megreya, A. M., & Burton, A. M. (2008). Matching faces to photographs: Poor performance in eyewitness memory (without the memory). *Journal of Experimental Psychology: Applied*, *14*(4), 364.

Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law*, *7*(1), 3-35.

O'Toole, A. J., Deffenbacher, K., Abdi, H., & Bartlett, J. C. (1991). Simulating the 'other-race effect' as a problem in perceptual learning. *Connection Science*, 3(2), 163-178.

O'Toole, A. J., Deffenbacher, K. A., Valentin, D., & Abdi, H. (1994). Structural aspects of face recognition and the other-race effect. *Memory & Cognition*, *22*(2), 208-224.

Pezdek, K., Blandon-Gitlin, I., & Moore, C. (2003). Children's face recognition memory: More evidence for the cross-race effect. *Journal of Applied Psychology*, *88*(4), 760.

Rhodes, G., Ewing, L., Hayward, W. G., Maurer, D., Mondloch, C. J., & Tanaka, J. W. (2009). Contact and other-race effects in configural and component processing of faces. *British Journal of Psychology*, *100*(4), 717-728.

Robertson, D., Black, J., Chamberlain, B., Megreya, A. M., & Davis, J. P. (2020). Super-recognisers show an advantage for other race face identification *Applied Cognitive Psychology*, *34*(1), 205-216.

Russell, R., Duchaine, B., & Nakayama, K. (2009). Super-recognizers: People with extraordinary face recognition ability. *Psychonomic Bulletin & Review*, *16*(2), 252-257.

Sangrigoli, S., & De Schonen, S. (2004). Recognition of own-race and other-race faces by three-month-old infants. *Journal of Child Psychology and Psychiatry, 45*(7), 1219-1227.

Sangrigoli, S., Pallier, C., Argenti, A. M., Ventureyra, V. A. G., & de Schonen, S. (2005). Reversibility of the other-race effect in face recognition during childhood. *Psychological Science*, *16*(6), 440-444.

Shah, P., Sowden, S., Gaule, A., Catmur, C., & Bird, G. (2015). The 20 item prosopagnosia index (PI20): relationship with the Glasgow face-matching test. *Royal Society Open Science*, *2*(11), 150305.

Sun, G., Zhang, G., Yang, Y., Bentin, S., & Zhao, L. (2014). Mapping the time course of other-race face classification advantage: a cross-race ERP study. *Brain Topography*, *27*(5), 663-671.

Tanaka, J. W., & Curran, T. (2001). A neural basis for expert object recognition. *Psychological Science*, *12*(1), 43-47.

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. The Quarterly Journal of Experimental Psychology Section A, 46(2), 225-245.

Tanaka, J.W., Heptonstall, B., & Campbell, A. (2019). Part and whole face representations in immediate and long-term memory. *Vision Research, 164,* 53 – 61.

Tanaka, J. W., & Taylor, M. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology*, *23*(3), 457-482.

Towler, A., White, D., & Kemp, R. I. (2017). Evaluating the feature comparison strategy for forensic face identification. *Journal of Experimental Psychology: Applied*, *23*(1), 47–58.

Verhallen, R. J., Bosten, J. M., Goodbourn, P. T., Lawrance-Owen, A. J., Bargary, G., & Mollon, J. D. (2017). General and specific factors in the processing of faces. *Vision Research*, *141*, 217-227.

Wan, L., Crookes, K., Dawel, A., Pidcock, M., Hall, A., & McKone, E. (2017). Face-blind for other-race faces: Individual differences in other-race recognition impairments. *Journal of Experimental Psychology: General*, *146*(1), 102.

Wilhelm, O., Herzmann, G., Kunina, O., Danthiir, V., Schacht, A., & Sommer, W. (2010). Individual differences in perceiving and recognizing faces—One element of social cognition. *Journal of Personality and Social Psychology*, *99*(3), 530.

Young, S. G., & Hugenberg, K. (2012). Individuation motivation and face experience can operate jointly to produce the own-race bias. *Social Psychological and Personality Science*, *3*(1), 80-87.

Young, S. G., Hugenberg, K., Bernstein, M. J., & Sacco, D. F. (2012). Perception and motivation in face recognition: A critical review of theories of the cross-race effect. *Personality and Social Psychology Review*, *16*(2), 116–142.

Figure Captions

Figure 1. Receiver operating characteristic (ROC) curves as a function of face race and participants' general face-processing ability. Each curve presents the average hit rate and false alarm rate for the most high-confidence decision (a judgement of "definitely yes" rather than "probably yes") in the lower left corner. Decisions of diminishing confidence ("probably yes" versus "maybe yes"; "maybe yes" versus "maybe no") are represented by the points that curve up and to the right. Curves closer to the diagonal identity line represent lower accuracy. Curves that deviate from the diagonal line in a more pronounced manner represent greater sensitivity.

Figure 2. Receiver operating characteristic (ROC) curves as a function of face race and participants' memory-based face-processing ability, measured by the CFMT+.

Figure 3a. Receiver operating characteristic (ROC) curves as a function of face race and participants' perceptual face-processing ability, measured by the GFMT.

Figure 3b. Sensitivity as a function of face race, cross-race contact, and participants' perceptual face-processing ability, measured by the GFMT.

Figure 4. Receiver operating characteristic (ROC) curves as a function of face race and participants' super-recognizer status.

Figure 5. Receiver operating characteristic (ROC) curves as a function of face race and participants race.

Figure 1

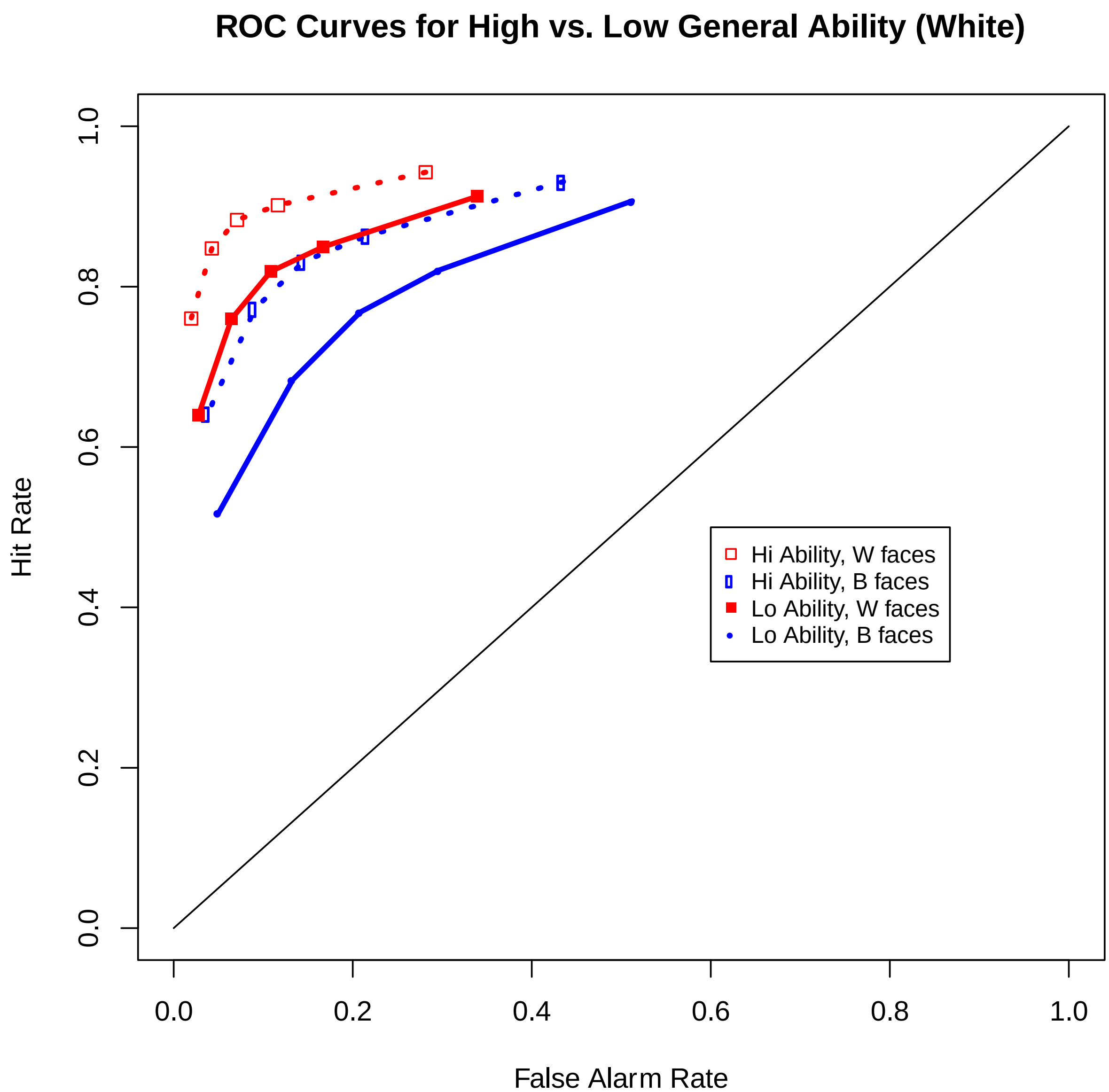

ROC Curves for High vs. Low General Ability (White)
Hit Rate
False Alarm Rate
0.0
0.2
0.4
0.6
0.8
1.0
Hi Ability, W faces
Hi Ability, B faces
Lo Ability, W faces
Lo Ability, B faces

Figure 2

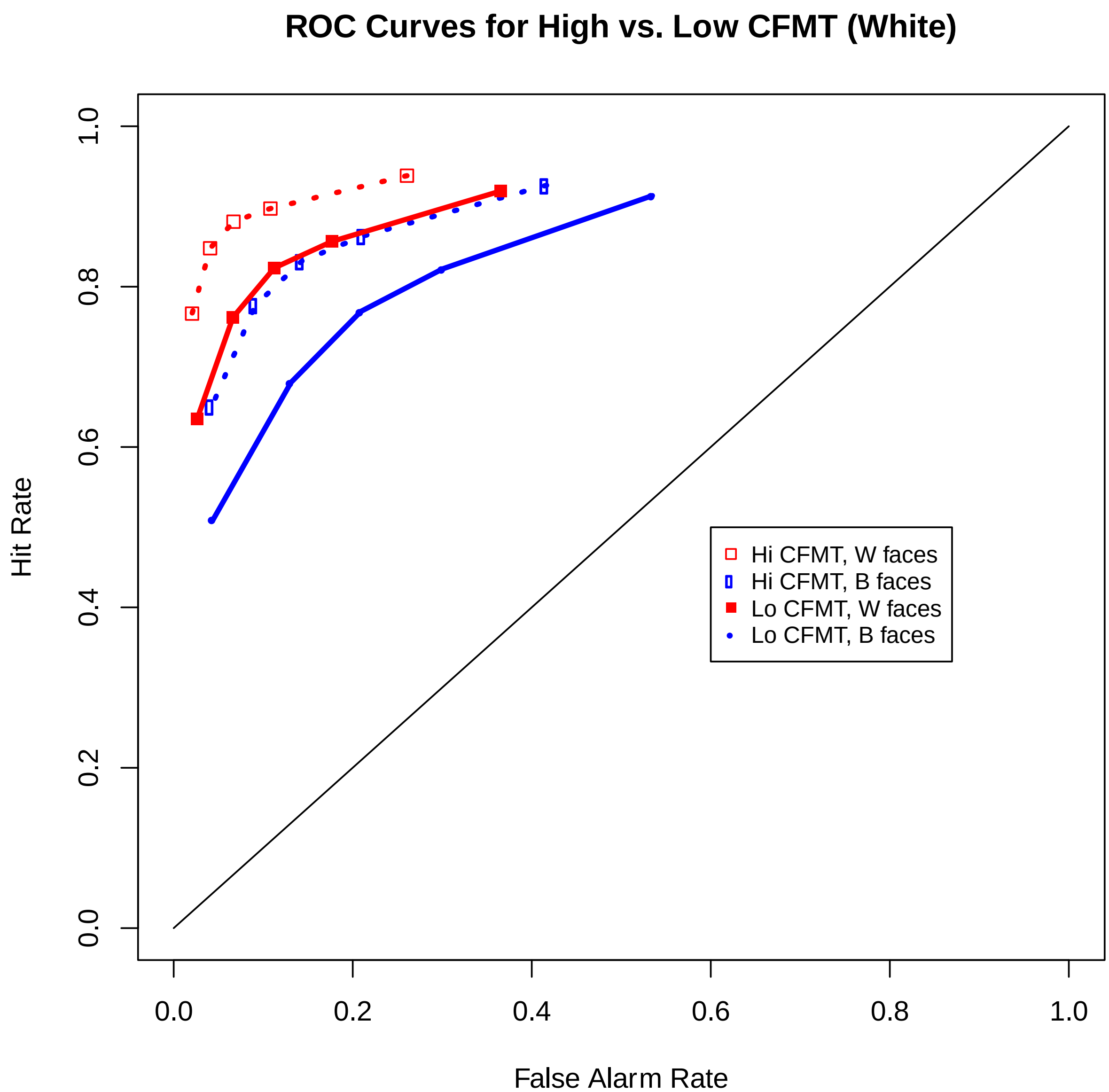
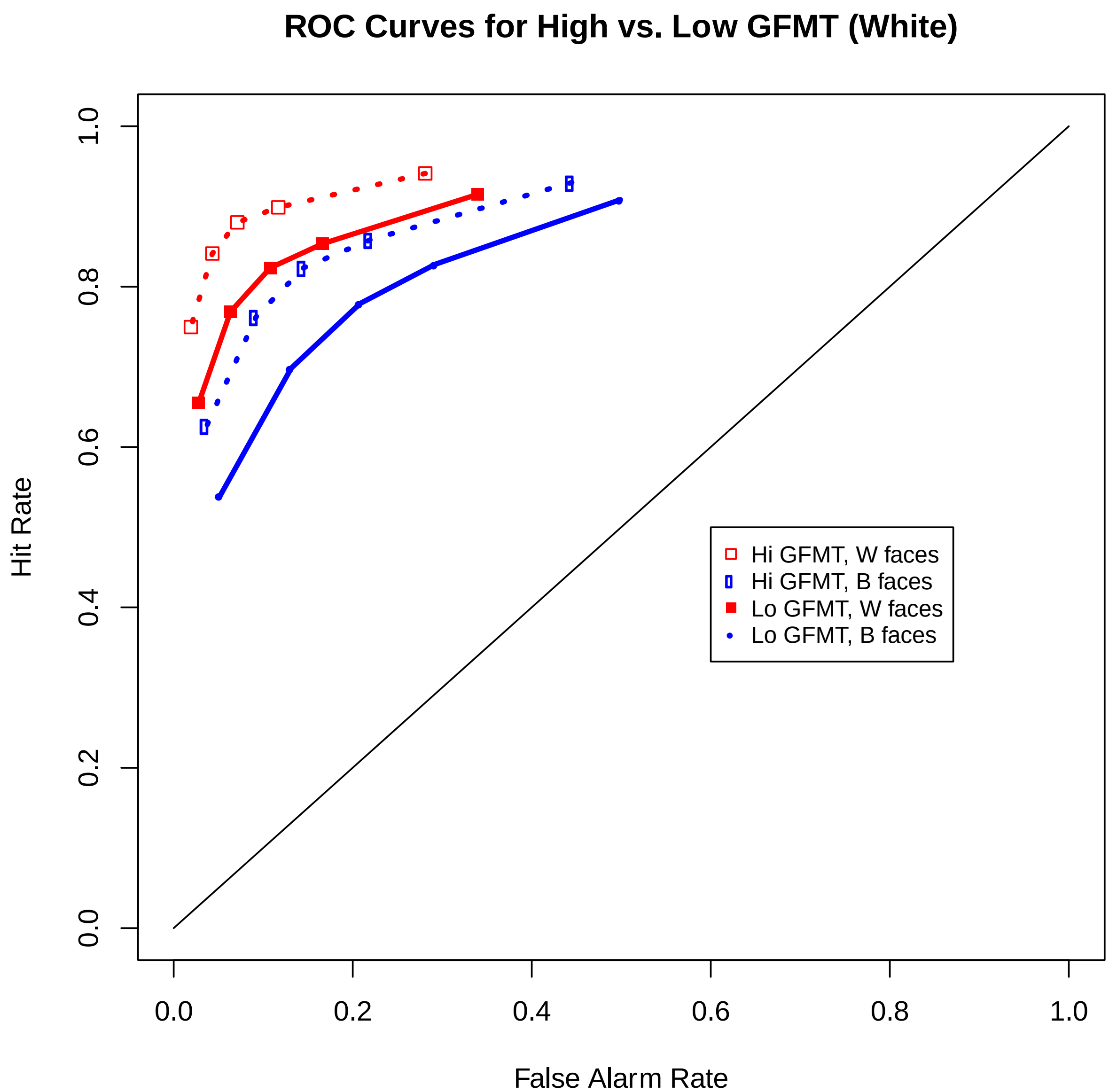

ROC Curves for High vs. Low CFMT (White)
Hit Rate
False Alarm Rate
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0
Hi CFMT, W faces
Hi CFMT, B faces
Lo CFMT, W faces
Lo CFMT, B faces

Figure 3a



ROC Curves for High vs. Low GFMT (White)
Hit Rate
False Alarm Rate
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0
Hi GFMT, W faces
Hi GFMT, B faces
Lo GFMT, W faces
Lo GFMT, B faces

Figure 3b

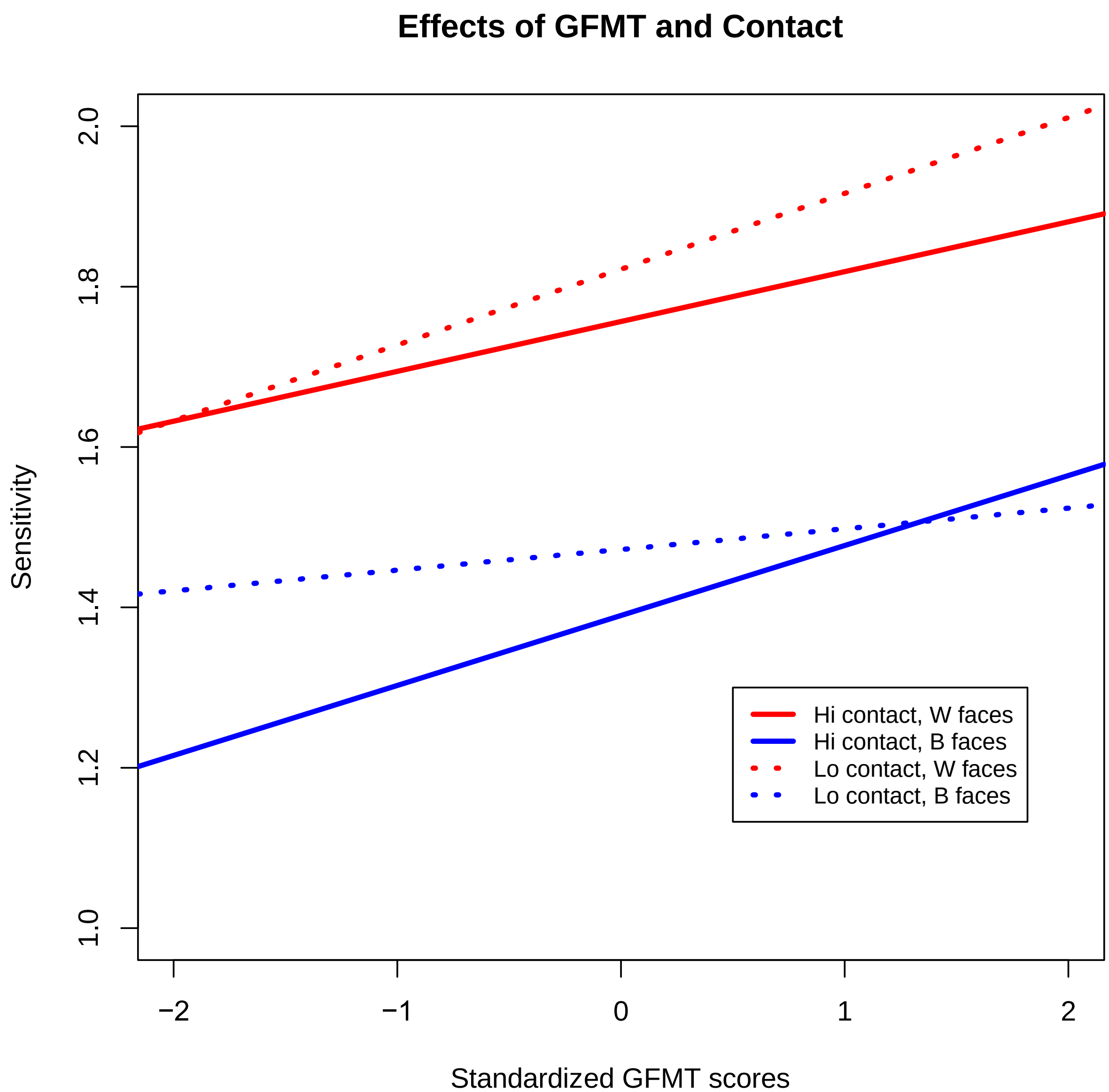
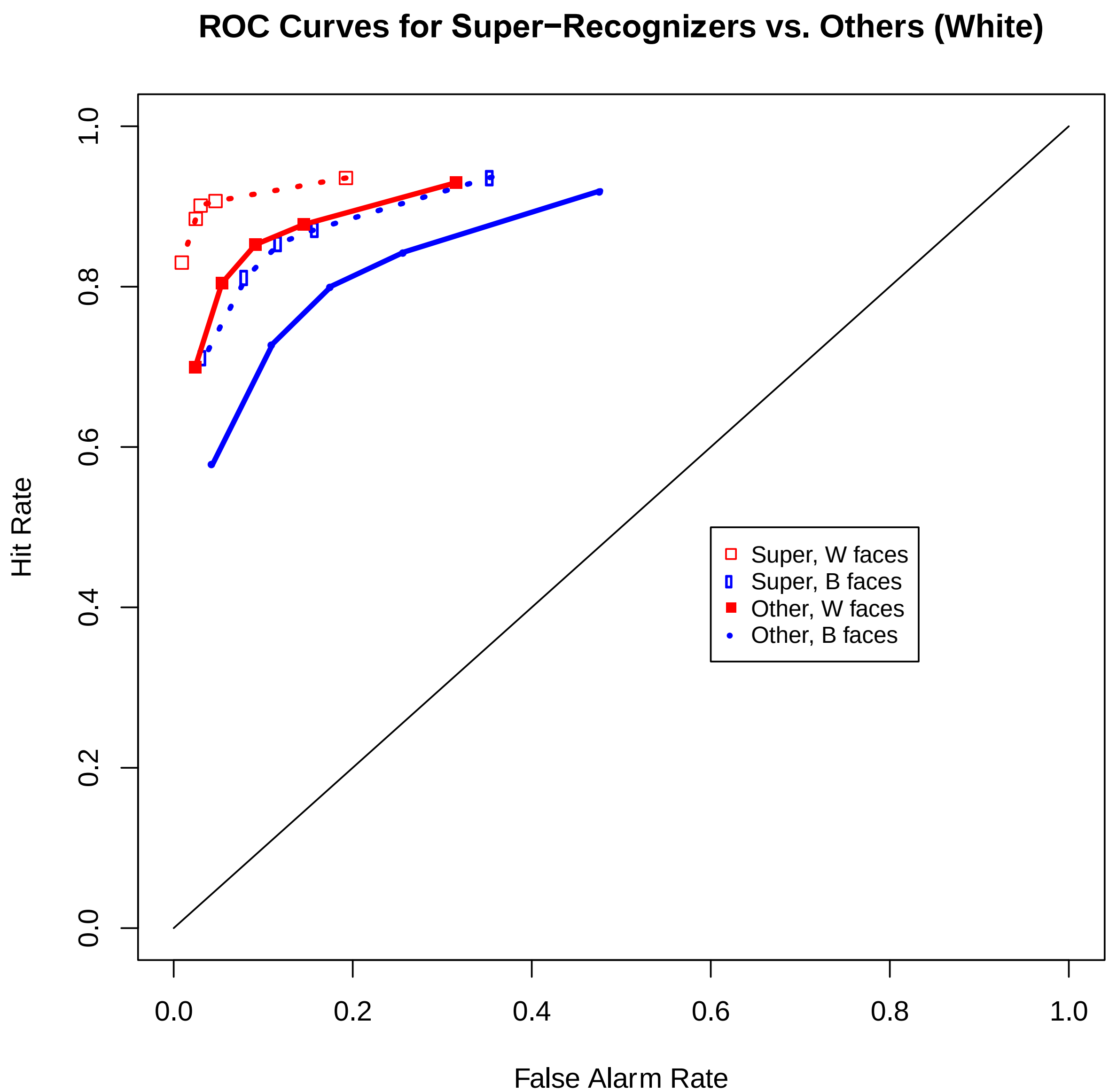
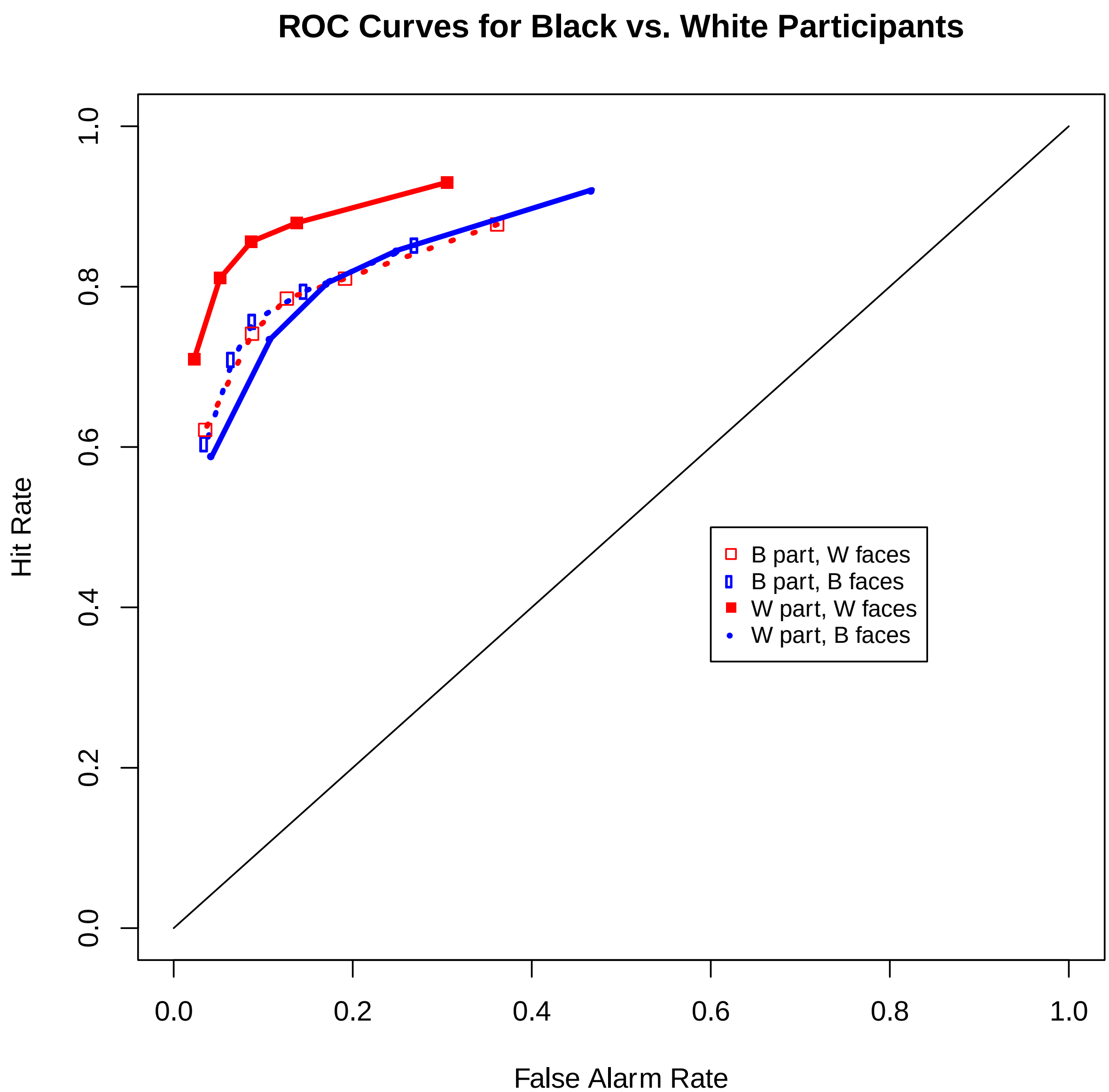

Effects of GFMT and Contact
Sensitivity
2.0
1.8
1.6
1.4
1.2
1.0
−2
−1
0
1
2
Standardized GFMT scores
Hi contact, W faces
Hi contact, B faces
Lo contact, W faces
Lo contact, B faces

Figure 4



ROC Curves for Super−Recognizers vs. Others (White)
Hit Rate
False Alarm Rate
0.0
0.2
0.4
0.6
0.8
1.0
Super, W faces
Super, B faces
Other, W faces
Other, B faces

Figure 5


ROC Curves for Black vs. White Participants
Hit Rate
1.0
0.8
0.6
0.4
0.2
0.0
B part, W faces
B part, B faces
W part, W faces
W part, B faces
0.0
0.2
0.4
0.6
0.8
1.0
False Alarm Rate